April 3, 2025

VIA ECF

The Honorable Ona T. Wang
Daniel Patrick Moynihan United States Courthouse
500 Pearl St.
New York, NY 10007

Re: *Authors Guild v. OpenAI Inc.,* No. 1:23-cv-8292-SHS-OTW (S.D.N.Y.); *Alter v. OpenAI Inc.*, No. 1:23-cv-10211-SHS-OTW (S.D.N.Y.)

Dear Magistrate Judge Wang:

Class Plaintiffs' RFP 62 seeks broad financial information for any Microsoft product that in any way uses one of OpenAI's large language models,[1] including the entire Microsoft Office suite of products. Microsoft has already agreed to produce (1) revenue and profit information for the products identified in the Complaint, as well as (2) revenue specifically tied to the licensing of LLM capabilities for the models at issue. This agreement covers all relevant revenue even under the most attenuated theories of causation. The request for revenue from the entire Microsoft Office suite of products (including all revenue from Word, Excel, and PowerPoint) is not reasonably related to the alleged infringement in this case. To the contrary, this is Class Plaintiffs' latest improper attempt to obtain discovery into company-wide Microsoft revenue.

The revenue information Plaintiffs seek has no causal relationship to the infringement by training allegations that Plaintiffs assert. Under 17 U.S.C. § 504(b), a plaintiff must establish defendant's "gross revenue" that is "reasonably related to the infringement." *On Davis v. The Gap, Inc.*, 246 F.3d 152, 160-61 (2d Cir. 2001) (gross revenue means "gross revenue reasonably related to the infringement, not unrelated revenues"). Courts in the Second Circuit have emphasized that the plaintiff's threshold burden is to demonstrate a "causal nexus" between the alleged infringement and the revenue in question. *Complex Sys., Inc. v. ABN Ambro Bank N.V.*, No. 08-cv-7497, 2013 WL 5970065, at *3-5 (S.D.N.Y. Nov. 8, 2013); *see On Davis*, 246 F.3d at 160–61 (plaintiff failed to meet their causal nexus burden when presenting revenue from parent company without any causal nexus to the Gap or the Gap ad at issue in the case); *Taylor v. Meirick*, 712 F.2d 1112 (7th Cir. 1983) ("If General Motors were to steal your copyright and put it in a sales brochure, you could not just put a copy of General Motors' corporate income tax return in the record and rest your case for an award of infringer's profits"). That rule applies equally to discovery. *See Structured Asset Sales, LLC v. Sheeran*, 433 F. Supp. 3d 608, 612 (S.D.N.Y. 2020) (rejecting request for revenue where no plausible factual basis to tie the infringing live performances of copyrighted song to non-concert revenue and merchandising); *see also Kaseberg v. Conaco, LLC*, 2016 WL 3997600, at *7–8 (S.D. Cal. July 26, 2016) (requiring production only

---

[1] Plaintiffs' motion focuses on RFP 62, but Microsoft has consistently applied the same relevance limitation to the related requests seeking financial information (RFP Nos. 63 and 65 and Interrogatory No. 6). The resolution of this dispute as to RFP 62 should therefore govern those requests as well.

Honorable Ona T. Wang   - 2 -   April 3, 2025

of the gross revenue from the specific episodes where allegedly infringed jokes appeared, not all revenue from the show). RFP 62 vastly exceeds that principle by demanding information on any Microsoft product that incorporates an OpenAI LLM, without regard to whether that product's revenue has a causal nexus to the act of using a book to train an LLM. This is akin to improperly seeking all of Microsoft's company-wide revenue.[2]

Although Class Plaintiffs claim that "profits derived from [products with integrations of allegedly infringing LLMs] are direct profits," they do not allege that a copy of any book in suit is incorporated into any Microsoft product. And indeed, Plaintiffs cite only to cases where the products contained a copy of the work at issue. *Bergt v. McDougal Littell*, 661 F. Supp. 2d 916, 927 (N.D. Ill. 2009) (defining profits from sales of textbook, the "product containing the copyrighted work," as direct because infringing painting "forms an integrated component of the textbook."); *Semerdjian v. McDougal Littell*, 641 F. Supp. 2d 233, 238, 248 (S.D.N.Y. 2009) (publisher produced 1.3 million copies of a literature textbook containing a reproduction of plaintiff's painting). To disgorge Microsoft's profits across any of its product lines incorporating LLM capabilities, all of the following would need to be true: (1) Author's book (i.e., work-in-suit) is copied and tokenized; (2) the tokenized data derived from the copy of Author's book is mixed with the mass of trillions of tokens used to train the LLM; (3) somehow that specific miniscule portion of the tokens represented by the copy of Author's book provides measurable value in the operation of the LLM; (4) the hypothetical specific value in the LLM attributable to this tiny portion of tokens is actually relied upon by the Microsoft product; and (5) customers care enough about the potential activation of the hypothetical value provided by the miniscule number of tokens that they purchase the Microsoft product. Merely to state this chain of hypotheticals is to refute it. *See Business Trends Analysts, Inc. v. Freedonia Group, Inc.*, 887 F.2d 399 (2d Cir. 1989) (alleged increase in value not recoverable as profits where not demonstrably attributable to the infringement).

Despite the vast overbreadth of Plaintiffs' request, Microsoft did **not** flatly refuse to provide any financial discovery. Rather, Microsoft offered to provide financial data regarding products where there is pricing specifically tied to the AI models at issue. In other words, Microsoft agreed to produce information where only assumption (5) in the hypothetical above could be met because the revenue dollars are tied to the use of LLMs. In trying to resolve this issue, Microsoft has agreed to ignore all the other assumptions above that would need to be met for this data to be relevant, such as that there is no allegation that any work in suit that may have been swirled into a giant mass of training material ultimately caused that revenue to be earned. Accordingly, Microsoft has agreed to produce revenue relating to the products Plaintiffs identified in their Complaint: Consumer Copilot (formerly known as Bing Chat), Copilot Pro (paid version of Consumer Copilot), Bing Chat Enterprise, Microsoft 365 Copilot (generative AI features in Microsoft 365/Office), Azure OpenAI Service, and GitHub Copilot. Microsoft's proffered production is generously more than sufficient to provide Plaintiffs with any revenue they can possibly argue meets the requisite causal connection test.

---

[2] Plaintiffs actually served a request for company-wide financials in RFP 64 (seeking "Documents sufficient to show Your balance sheet, income statement, and cash flow statement, on a monthly basis during the relevant time period."). This letter motion represents Plaintiffs' attempt to circumvent the causal nexus requirement and seek this overbroad and irrelevant company-wide financial information through RFP Nos. 62, 63, and 65 and Interrogatory No. 6.

Honorable Ona T. Wang   - 3 -   April 3, 2025

   This dispute is easily distinguishable from the out-of-circuit scenarios cited by Plaintiffs. In *Prolo v. Blackmon*, a California court compelled financial discovery where the defendant had flatly refused to produce discovery on any revenue related to an infringing photograph, claiming no revenues were attributable to use of the copyrighted photograph in product/performance promotions. 2022 WL 2189643, at *9 (C.D. Cal. Mar. 25, 2022). Here, Microsoft has already agreed to produce financial information for products that incorporate the accused technology, despite that Plaintiffs do not allege that any asserted copyrighted works are included or in use in any Microsoft product or product promotion. In *Systems America, Inc. v. Rockwell Software, Inc.*, Systems America similarly introduced "unrebutted evidence" that Rockwell touted its ability to offer allegedly misappropriated source code and trade secrets "in connection with its sales and marketing of all the products in issue." 2007 WL 1593219, at *2 (N.D. Cal. June 1, 2007). Microsoft is neither offering Plaintiffs' works for sale nor using Plaintiffs' works to promote its products in any fashion.

   Plaintiffs' Description of Microsoft's Limits is Both Factually and Procedurally Incorrect. While disputing the relevance of models that post-date GPT 3.5 due to a lack of evidence that any later model was trained on Plaintiffs' works, Microsoft agreed in meet and confer to produce revenue for products using GPT-3, GPT 3.5, GPT 3.5 Turbo, GPT-4, and GPT-4 Turbo models. In keeping with this Court's prior guidance, Microsoft does not agree to produce revenue relating to GPT-4o and GPT-mini and has consistently taken that position across all discovery. *See* ECF 293 at 2 (denying Class Plaintiffs' motion at ECF 270), 5 (denying News Plaintiffs' motion at ECF 287, 279). As this Court recognized in denying discovery into future or later products not named in the complaint, Plaintiffs are not entitled to discovery into every aspect of Microsoft's business. *Id.* at 5.

   Class Plaintiffs mischaracterize this Court's prior ruling regarding the LLMs subject to discovery, as this Court never decided whether the LLMs upon which OpenAI voluntarily agreed to produce discovery were within the appropriate scope of discovery. *See id.* at 2. Indeed, in responding to that motion, OpenAI made clear that some of the models were not properly within the scope. *See* ECF 281 at 1 ("This lawsuit, filed by book authors, centers around specific books datasets that were used to train GPT-3 and 3.5. OpenAI agreed to provide discovery into these and a number of *other* models—including GPT-1, GPT-2, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo—and, beyond that, to supplement its response to Interrogatory 11 . . .") (cleaned up). And the fact that a separate Defendant agreed to compromise in a prior dispute has no bearing on Microsoft's position and cannot be used to circumvent what this Court has already concluded— the proper focus is on the products and models ***presently at issue*** in the case.

   For the foregoing reasons, Microsoft respectfully requests that the Court deny Plaintiffs' letter-motion to compel further responses to RFP No. 62 (and by extension RFP Nos. 63 and 65 and Interrogatory No. 6, which involve the same issue).

 Respectfully submitted,   Respectfully submitted,

 /s/ Annette L. Hurst   */s/ Jared B. Briant*
 Annette L. Hurst   Jared B. Briant

Counsel for Defendant Microsoft Corporation